

Weekly Report

2013-11-10

Feiran Wu

1 The basic research route of multivariate time series

Our goal is to transfer the raw, pure numerical multivariate time series into time event model (event abstract). Then visualize this updated model.

So the basic task can be divided as:

T1. Extract the event (motif, burst, cycle).

T2. Visualize these events.

Now we concentrate to the T1.

2 The general route and progress in event extracting

As figure 1 (a) shown, there are three levels in multivariate time series. I organize the general idea as Figure 2 (b) depicted. The block in red is the methods we have tried (The detail can be seen in Guo Fangzhou's weekly report in this week). The block in green is the way we plan to try in the next week.

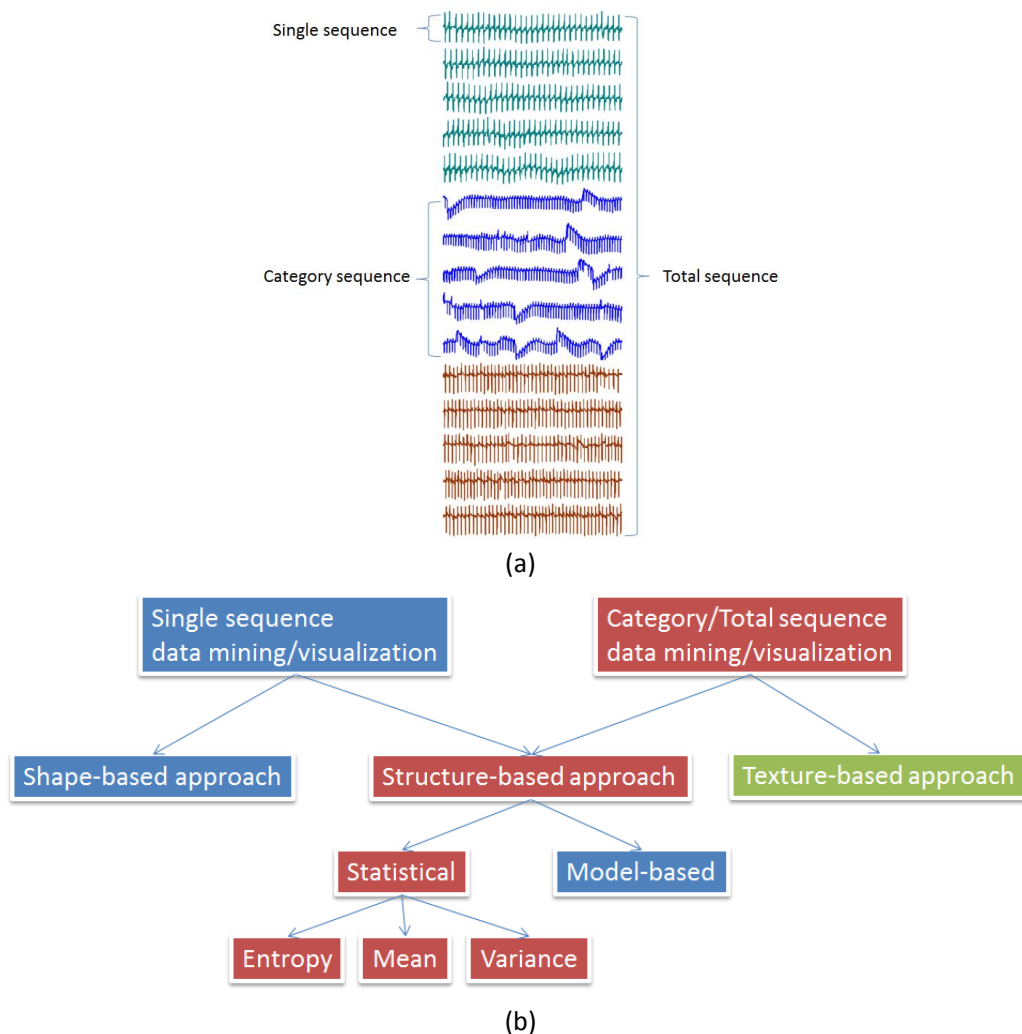


Figure1 (a) Three data levels of time-series. (b) The general idea block diagram.

3 Our idea in detail

The most popular ways in pattern extracting of time series, either data mining or visualization, can be categorized to the single sequence approach. Some visualization system concentrated to the multivariate time series visual design as the traffic jam analysis made by Yuan's group.

But the time-series is too long to glance and the occurrence events should be observed in multivariate. So how to locate these events? We want the statistical method as the pre-select method. As last week I reported, one of the statistical values - the entropy, can locate some events. After we locate one event, then I tend to use the texture-based approach to find the similar patterns (motifs).

If we sell our idea as an analysis pipeline/process, There are * steps as follows:

Step 1. Raw data import.

Step 2. Select the time sequences/attributes. Reorder them.

Step 3. Select a range and moment according to the statistical values (e.g. Entropy).

Step 4. Texture-based approach to find Top-K similar objects.

Step 5. Accept or reject. Make a choice.

If we sell our idea as a visual analysis model. We should place an emphasis on dealing every situation at Step 2 and Step 3. The general idea is to define a formal language which is convenient to express the matrix operations. We consider the total the pixel-like visual design of multivariate time series as a $n*m$ matrix K , where n is the number of attributes/variates and m is the length of all the time series. From the step 3 The submatrix K' selected. The step 4 is to find the Top-K similarities of K' .

4 Our jobs in the next week

To implement the texture-based method CK-distance (Keogh 2010) which is said better than SIFT, filter-band and the other approach in texture match (Keogh 2013).